ABSTRACT OF THE DISCLOSURE

The use for biological screening purposes of a subset (library) of a large combinatorially accessible chemical universe increases the efficiency of the screening process only if the subset contains members representative of the total diversity of the universe. In order to insure inclusion in the subset of molecules representing the total diversity of the universe under consideration, valid molecular descriptors which quantitatively reflect the diversity of the molecules in the universe are required. A unique validation method is used to examine both a new three dimensional steric metric and some prior art metrics. With this method, the relative usefulness/validity of individual metrics can be ascertained from their application to randomly selected literature data sets. By the appropriate application of validated metrics, the method of this invention selects a subset of a combinatorial accessible chemical universe such that the molecules of the subset are representative of all the diversity present in the universe and yet do not contain multiple members which represent the same diversity (oversample). The use of the neighborhood definition of a validated metric may also be used to combine (without oversampling the same diversity) any number of combinatorial screening libraries.